

周报

- 1、本周前面几天大家都在调研 1) 数据整合中不同数据源的可信度问题
2) 如果处理数据整合中数据的不一致和模糊性问题。大家都看了相关的材料和论文。不过 palantir 中并没有对数据源或者数据的不一致做任何特殊的处理，目前这部门我们也不打算再很深入地继续研究下去，这部分就不多说了。
- 2、根据周四的讨论，本周的剩余时间大家分别在安排的任务：
 - 1) 马晓红：进行整体框架的搭建。打算利用 eclipse 的插件 EMF 来建构模型，目前一些简单基本的数据模型已经构建的差不多，争取下周构建好，并和小组其他成员进行商讨。
 - 2) 杨哲：实现 facebook、twitter 等社交网络的爬虫接口。
 - 3) 王琦：继续进行本体的构建。周末发现 <http://schema.org/> 是个很好的网站，这是一个已经成型的 scheme，并且是各个搜索引擎全力去推广的。该网站上已经提供了它里面的所有类的继承结构和每个类的属性，但是没有类之间的关系。不过我觉得其中的类的继承关系和属性已经整理的不错了，我们完全可以拿来直接用，然后再添加上一些其他的属性和类之间的关系。王琦会先研究一下这个东西，然后看应该如何使用。
 - 4) 李嘉华：有两个通用的语义数据集 DBpedia 和 YAGO，在探究这两个我们能否使用以及如何使用。

下周任务：

1. 接着本周未完成的工作下周继续进行。

Palantir 小组计划

	工作计划
第 一 周 (11.24-11.30)	1) 设计整个系统的数据模型、框架结构 (马晓红) 2) 所使用的几个数据源的调研 (李嘉华) 3) 数据源接口的实现 (杨哲) 4) 本体的构建及写相关文档 5) 调研数据源的可信度和不一致问题 (林明)
第 二 周 (12.01-12.07)	1) 系统中局部本体和全局本体的基本功能的实现 (李嘉华) 2) 数据源中查询的接口的实现 (杨哲) 3) 本体的构建及写相关文档 (王琦) 4) 设计整个系统的数据模型、框架结构 (马晓红) 5) 调研数据源的可信度和不一致问题 (林明)
第 三 周 (12.08-12.14)	1) 系统中全局查询与局部查询的映射 (马晓红) 2) 系统 gui 的实现包括显示图结构、选中节点可以显现详细信息, 提供界面上一个实体的关联实体的查询 (杨哲、王琦、林明) 3) 从各个局部数据源中查询结构的实体对其 (李嘉华)
第 四 周 (12.15-12.21)	1) 继续完成界面的功能 (杨哲、王琦、林明) 2) 前后端数据连通 (李嘉华)
第 五 周 (12.22-12.28)	3) 继续完成界面的功能 (杨哲、王琦、林明) 前后端数据连通 (李嘉华)